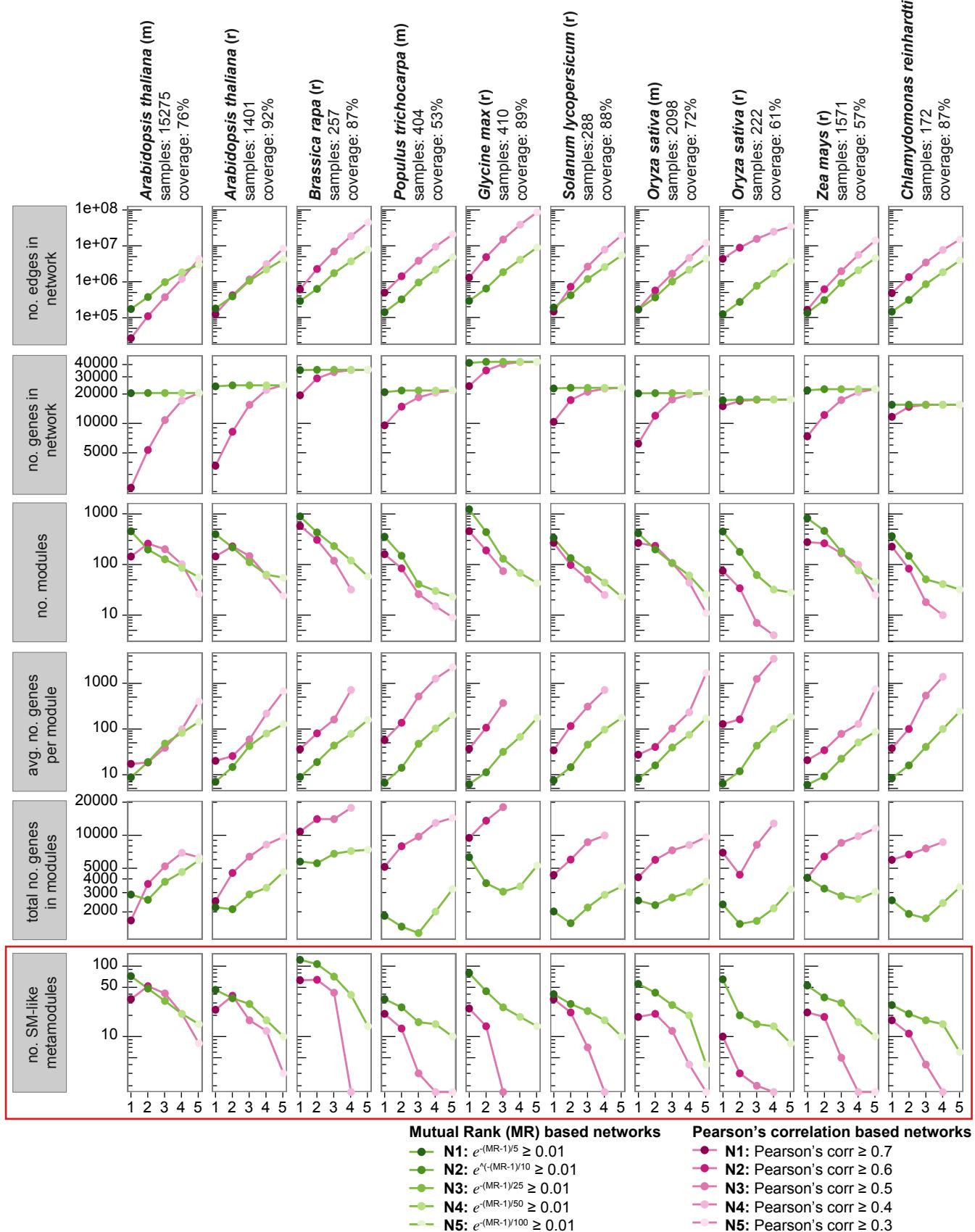
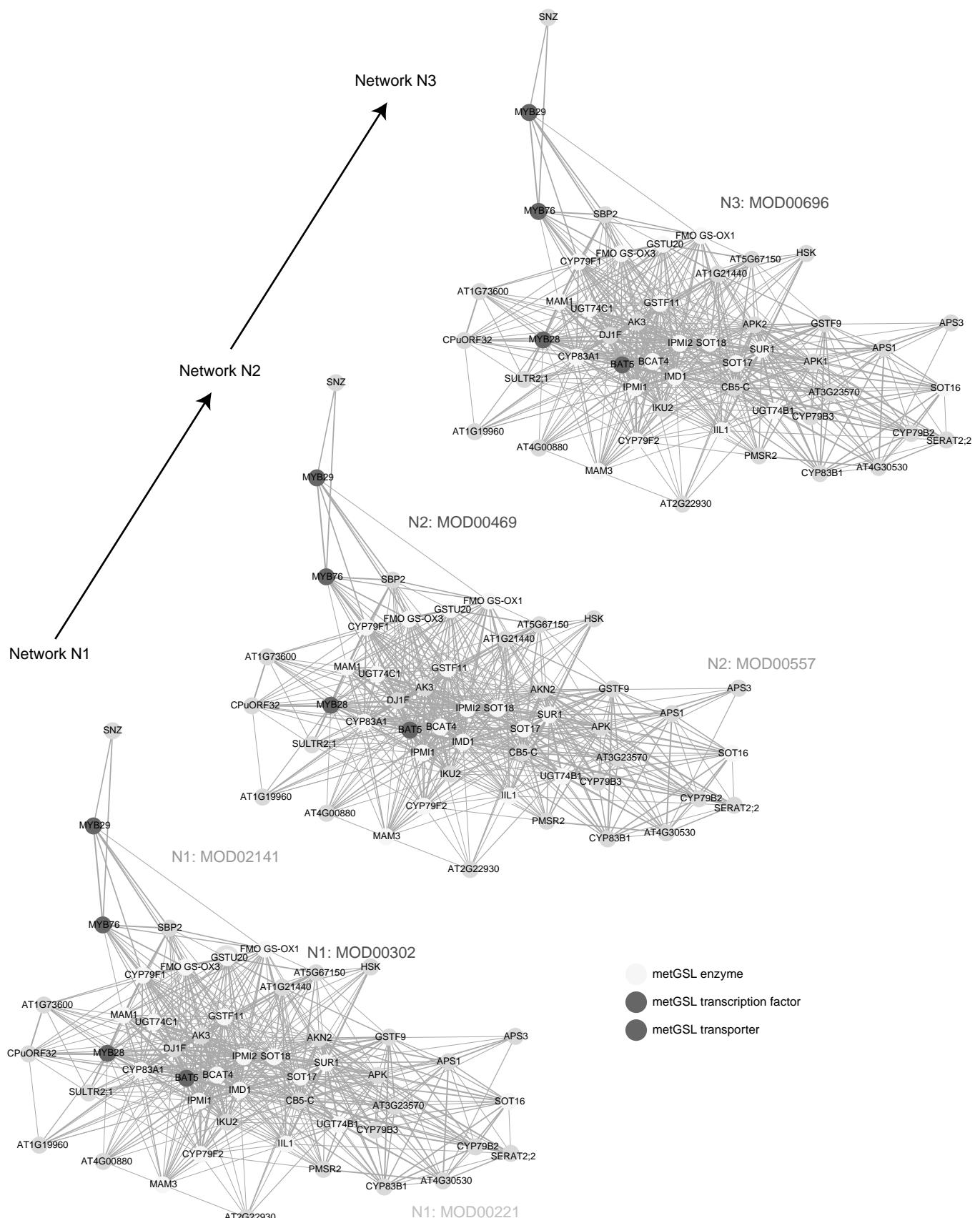


Supplemental Figure 1. MetaCyc pathway enrichment analysis of experimentally characterized genes in *A. thaliana*. The x axes (top) depict \log_{10} transformations of the relative change ratio (RCR) for genes with experimentally verified functions in modules versus those not assigned to modules. Significantly enriched ($RCR > 0$) or depleted ($RCR < 0$) MetaCyc pathway categories are indicated by the dark blue and black bars for secondary metabolism and all other categories, respectively; * $P \leq 0.05$, ** $P \leq 0.005$, *** $P \leq 0.0005$ (Benjamini & Hochberg adjusted P values, hypergeometric test; Table S4). Grey arrows indicate that the bottom 8 pathway categories are children of 'Biosynthesis' in the MetaCyc hierarchy. Red rectangle indicates the results from Mutual Rank network N1 (microarray dataset) presented in the main text (Figure 2A). (Supports Figure 2.)

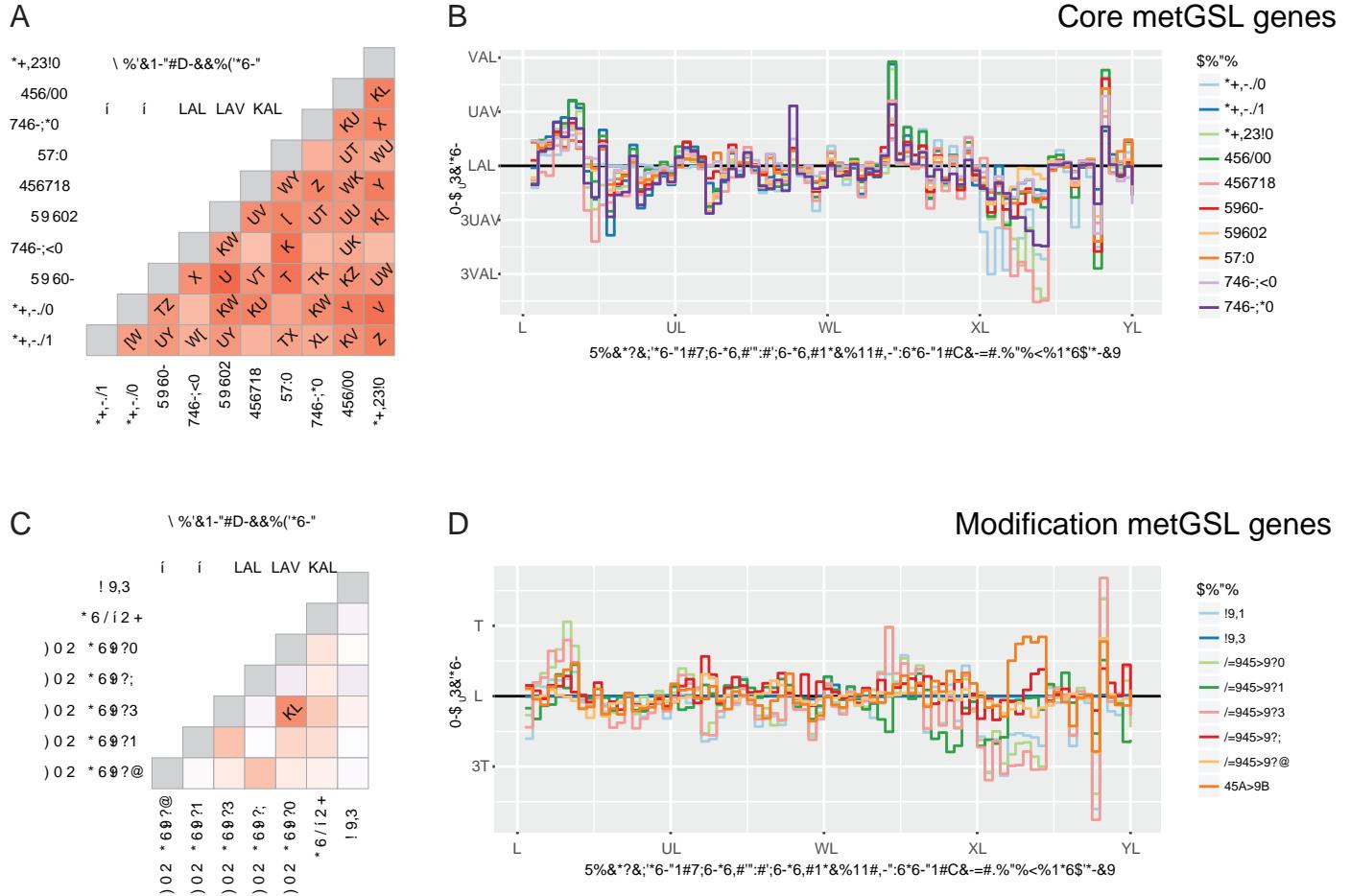


Supplemental Figure 2. Comparison of Mutual Rank (MR)-based and Pearson's correlation-based networks using 10 microarray (m) and RNAseq (r) co-expression datasets. MR scores and correlations were downloaded from ATTED-II and converted to network edge weights using the formulas at the bottom right; networks were ordered such that N1 was the smallest network with the fewest nodes/edges, and N5 was the largest network with the most nodes/edges. Modules of tightly co-expressed genes were detected by ClusterONE using default parameters (P value ≤ 0.1). Networks with missing values—e.g., Pearson's correlation-based N5 in *B. rapa* (r)—indicate that ClusterONE exceeded the 24-hour allowable runtime and did not finish; median runtime was 0.06 hours with access to 12 GB of RAM. SM-like modules were collapsed into meta modules of non-overlapping gene sets. Red rectangle indicates the results from MR-based networks N1-N3 presented in the main text (Figure 2B). Comparing the number of edges and genes (top two plots) in the microarray-based versus RNAseq-based networks in *A. thaliana* and *O. sativa* demonstrates how the Pearson's correlation-based networks (pink) were more variable in response to differences in the number of experimental samples and genome coverage in the co-expression analyses; in contrast, the MR-based networks (green) were less affected by these differences. (Supports Figure 2.)



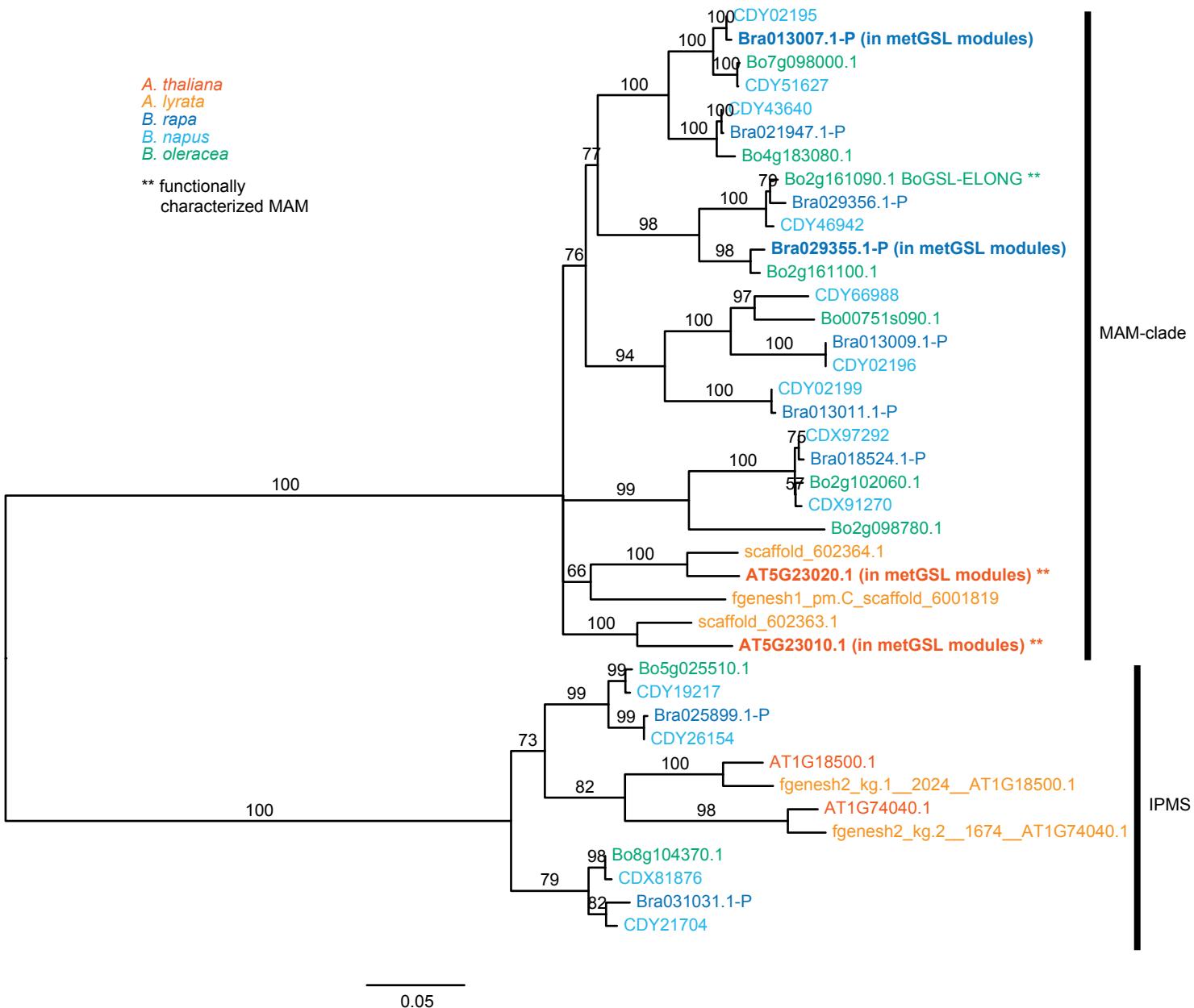
Supplemental Figure 3. Overlapping co-expressed modules recover the pathway for metGSL biosynthesis in *A. thaliana*. Nodes in the network maps represent genes. Edges connecting two genes represent the mean weight (transformed MR) for the association across all three networks (N1-N3). Network maps were drawn using a Fruchterman-Reingold force-directed layout using the igraph R package (<http://igraph.org>). Genes involved in metGSL biosynthesis are color-coded according to their function: orange for enzymatic genes, dark blue for those encoding transporters, and light blue for those encoding transcription factors. In the smallest network, N1, metGSL genes were assigned to three modules (MOD02141, MOD00302, and MOD00221). Note the overlapping nature of MOD02141 and MOD00302, with three genes (CYP79F1, FMO GS-OX3, and FMO GS-OX1) being assigned to both modules. In networks N2 and N3, the modules are fewer and larger, recovering more of the metGSL pathway into a single module at the expense of including additional genes (gray circles) that may or may not be involved in the pathway.

(Supports Figure 3.)



Supplemental Figure 4. Comparison of degree of gene co-expression in core versus terminal modification genes in metGSL biosynthesis. In general, the core GSL genes show strong co-expression (A) and behave similarly across perturbation experiments (B). Modification genes are not co-expressed (C) and are more variable across perturbation experiments (D), explaining why fewer of these genes are recovered in co-expression modules. Diagonal numbers in heatmaps indicate MR scores; squares are blank if MR > 100. The ATTED-II datasets used in this analysis consist only of pairwise correlations of gene co-expression. Data from individual experiments/conditions that underlie these correlations are not provided for bulk download. We used Genevestigator to visualize how gene expression within a known SM pathway varies across a similar dataset. Difference in gene expression between the control condition and all RNAseq perturbation experiments in wild-type genetic background in *A. thaliana* are plotted for both core and modification metGSL genes.

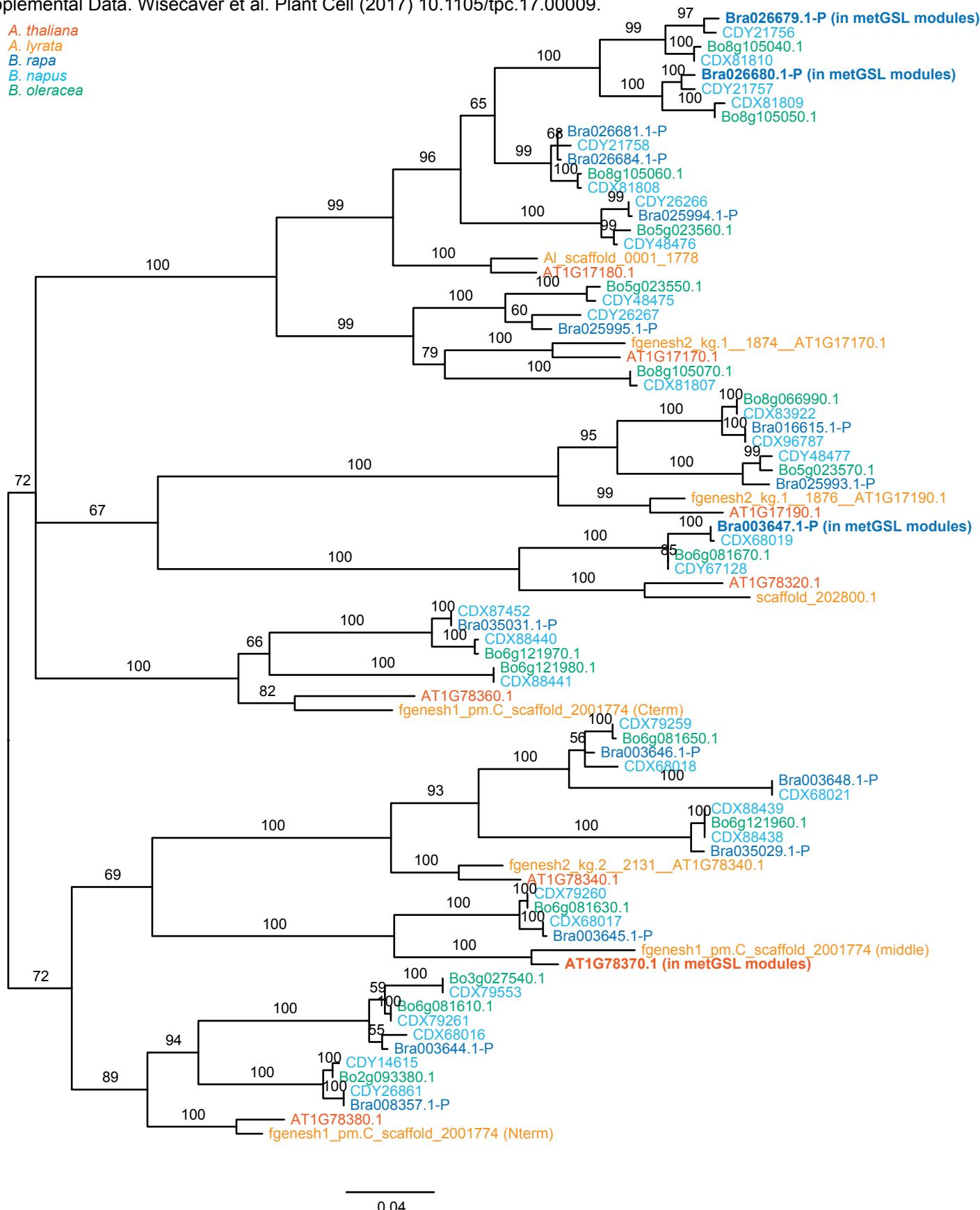
(Supports Figure 3.)



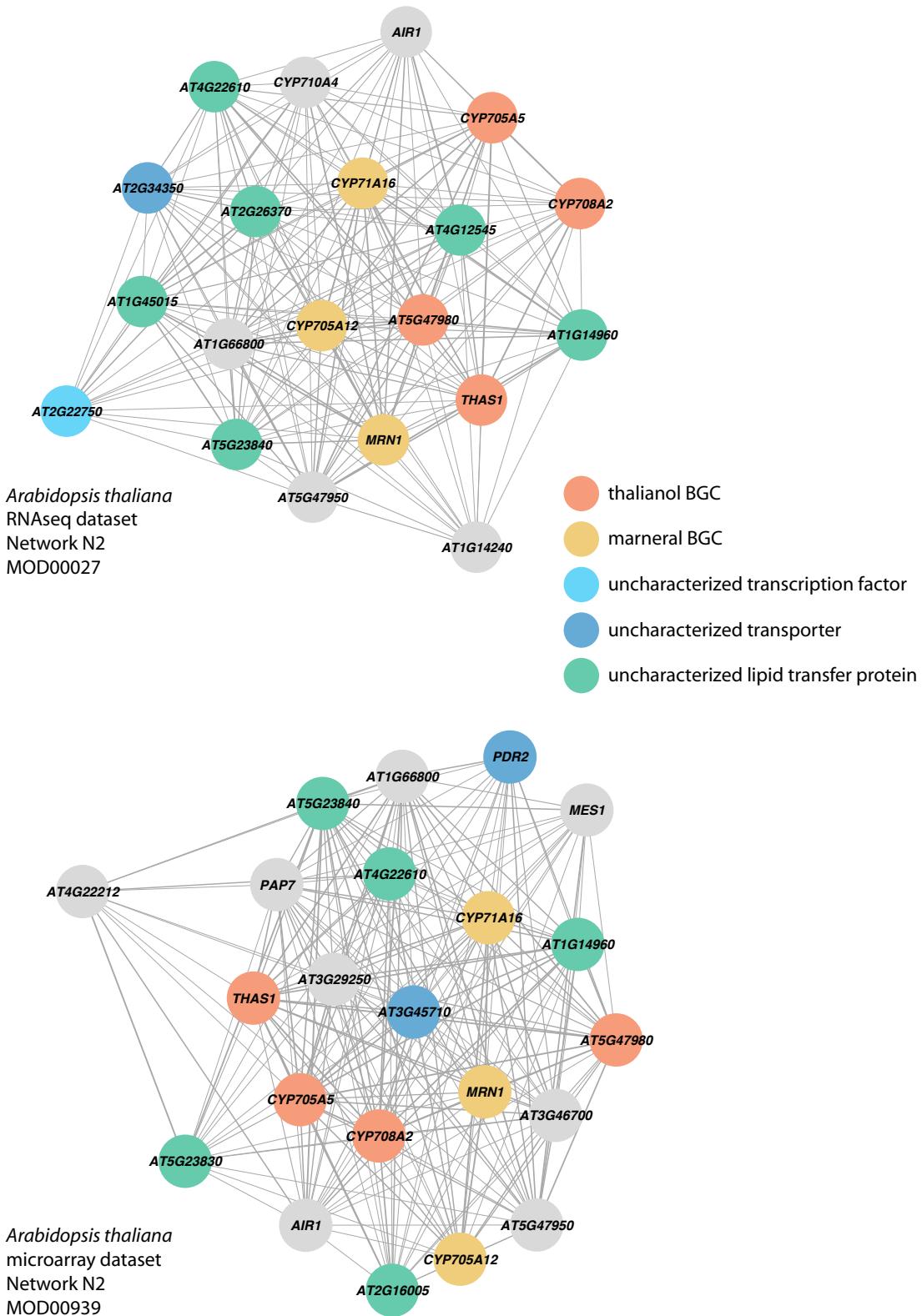
Supplemental Figure 5. Maximum likelihood phylogeny of Brassicaceae MAM and IPMS sequences. Tree is midpoint rooted. Branches with less than 50% bootstrap support were collapsed using TreeCollapserCL4. Numbers above branches indicate the percentage of bootstrap replicates supporting each branch. Scale bar indicates the mean expected rate of nucleotide substitution per site. MAMs from *A. thaliana* and *B. rapa* that are co-expressed with other metGSL genes are bolded. Note: this phylogeny provides no resolution for labeling *B. rapa* genes as either MAM1- or MAM3-like.

(Supports Figure 4.)

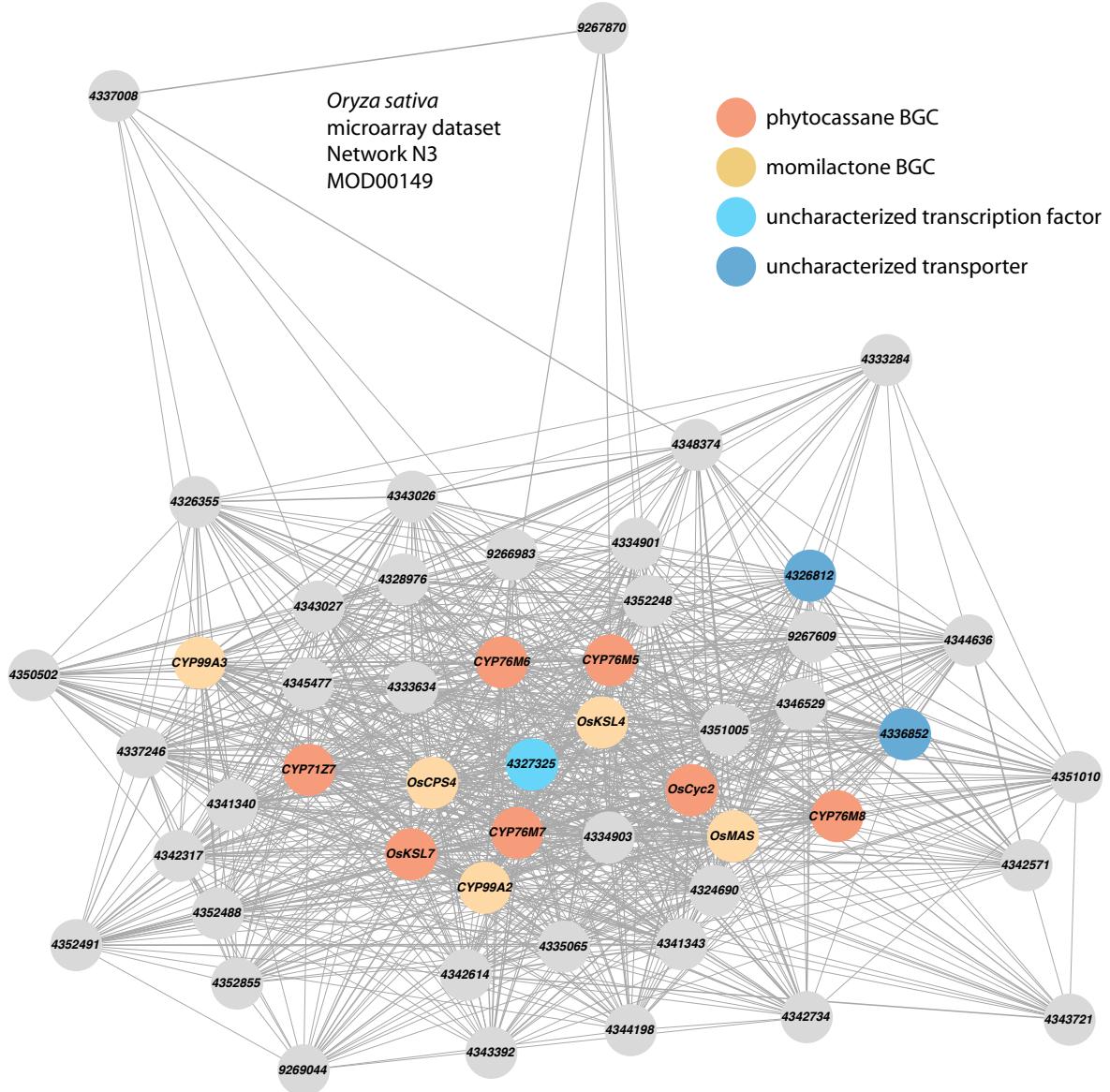
A. thaliana
A. lyrata
B. rapa
B. napus
B. oleracea



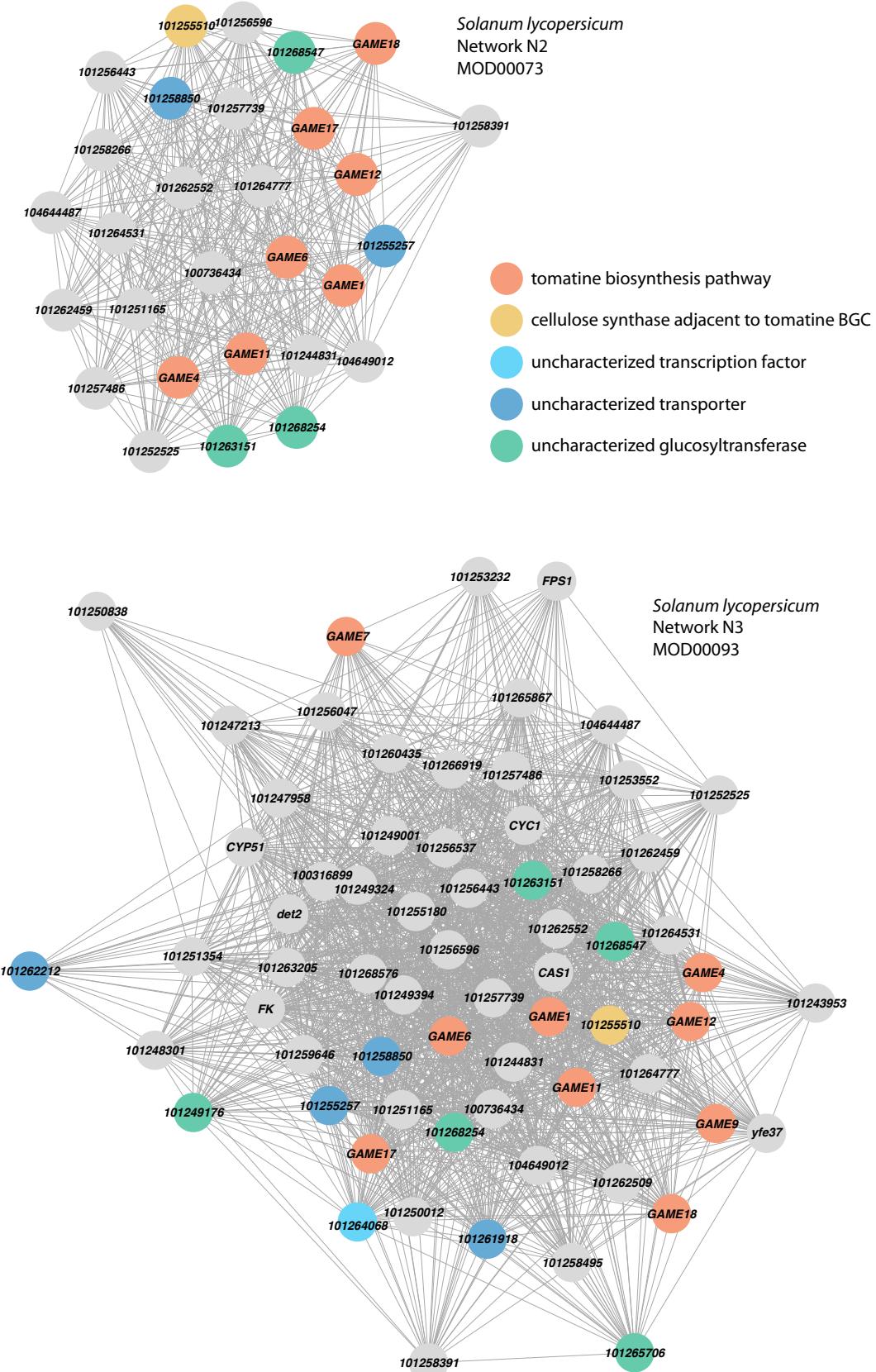
Supplemental Figure 6. Maximum likelihood phylogeny of Brassicaceae GSTU sequences. Tree is midpoint rooted. Branches with less than 50% bootstrap support were collapsed using TreeCollapserCL4. Numbers above branches indicate the percentage of bootstrap replicates supporting each branch. Scale bar indicates the mean expected rate of nucleotide substitution per site. *A. thaliana* GSTU sequences are labeled according to ensembl. The predicted CDS from *A. lyrata* fgenesh1_pm.C_scaffold_2001774 was three times as long as other GSTUs in this tree, consisting of three GSTU domain duplicates. Individual domains from this CDS were evaluated independently (see 'Nterm', 'middle', and 'Cterm' leaves) and each group with a separate GSTU clade in the phylogeny. Note: GSTUs from *A. thaliana* and *B. rapa* that are co-expressed with metGSL genes (bolded) are not orthologous. (Supports Figure 4.)



Supplemental Figure 7. Network maps of co-expression modules involved in thalianol and marnerol triterpenoid biosynthesis in *A. thaliana*. Nodes in the maps represent genes. Edges connecting two genes represent the weight (transformed MR score) for the association in the specified network. Network maps were drawn using a Fruchterman-Reingold force-directed layout using the igraph R package (<http://igraph.org>).
(Supports Figure 6.)



Supplemental Figure 8. Network map of co-expression module involved in momilactone and phytocasane diterpenoid biosynthesis in *O. sativa*. Network maps are drawn as described in Supplemental Figure 7. (Supports Figure 6.)



Supplemental Figure 9. Network maps of co-expression modules involved in tomatine biosynthesis in *S. lycopersicum*.
Network maps are drawn as described in Supplemental Figure 7.
(Supports Figure 6.)

%TAAIB@B"#!)R#!F)`\$GBJ!WB9)B#)!F)2!"#,BII)OD/0.P)0/F00/<a#AJF0.F///;F

